BEHAVIOUR & INFORMATION TECHNOLOGY, JANUARY–FEBRUARY 2004, VOL. 23, NO. 1, 53–64

Long-term working memory and interrupting messages in human – computer interaction

ANTTI OULASVIRTA†‡ and PERTTI SAARILUOMA§

†Helsinki Institute for Information Technology, Finland

‡Cognitive Science Unit, Department of Psychology, University of Helsinki, Finland

§Agora Human Technologies Center, University of Jyväskylä, Finland; e-mail: psa@cc.jyu.fi

Abstract. The extent to which memory for information content is reliable, trustworthy, and accurate is crucial in the information age. Being forced to divert attention to interrupting messages is common, however, and can cause memory loss. The memory effects of interrupting messages were investigated in three experiments. In Experiment 1, attending to an interrupting message decreased memory accuracy. Experiment 2, where four interrupting messages were used, replicated this result. In Experiment 3, an interrupting message was shown to be most disturbing when it was semantically very close to the main message. Drawing from a theory of long-term working memory it is argued that interrupting messages can both disrupt the active semantic elaboration of content during encoding and cause semantic interference upon retrieval. Properties of the interrupting message affect the extent and type of errors in remembering. Design implications are discussed.

1. Introduction

It is seldom noticed that there may be two different approaches to human – computer interaction (HCI). Firstly, one may be interested in modifying the *usability* of user interfaces according to some artefact design rationale. In this work, user-tests provide the essential way of applying psychological knowledge. Secondly, one may be interested in the nature of the psychological mechanisms involved in using computers. In this line of research, the main goal is to formulate psychological design rationales for designers. This approach could be termed *user psychology* in order to clarify the difference from the term usability research. Accordingly, we shall here investigate the role of an influential cognitive theory, called the theory of long-term working memory (Ericsson and Kintsch 1995), in human-computer interaction.

It is fairly well established today that human working memory can be divided into two functional components: *short-term working memory* (STWM) and *long-term working memory* (LTWM) (Ericsson and Kintsch 1995, Ericsson and Lehman 1996, Ericsson and Delaney 1999). Alternative terminology has also been used, but the basic idea is very similar (e.g. Richman *et al*. 1995, Gobet and Simon 1996, Gobet 1998). STWM is a capacity- and time-limited store, located in the frontal lobes, the function of which is in actively updating and manipulating representations, switching and dividing attention between tasks, selection of relevant information, and inhibition of irrelevant information (Baddeley 1986, 1996, 2000, Cowan 2001). LTWM can be seen as a kind of intermediate memory within the classic long-term memory. The LTWM theory supposes that (1) information is incidentally encoded to LTWM during skilled activities, (2) speed of encoding increases with practice, (3) encoding is meaningful in accord with a vast body of prior knowledge, (4) information is encoded into organized systems of retrieval cues called retrieval structures, (5) selectivity of encoding increases with practice in accord with task demands, and (6) during encoding, access to task-relevant information may be maintained by means of pointers residing in STWM that refer to retrieval structures in LTWM. A prerequisite for LTWM memory skills is several years of practice. Furthermore, when acquired, these memory skills are specific to the domain of expertise. (Saariluoma 1991, Ericsson and Kintsch 1995, Saariluoma 1995, Saariluoma and Kalakoski 1997, 1998).

Behaviour & Information Technology
ISSN 0144-929X print/ISSN 1362-3001 online © 2004 Taylor & Francis Ltd
http://www.tandf.co.uk/journals
DOI: 10.1080/01449290310001644859

Task interruption provides a practical HCI phenomenon in which LTWM is involved. In HCI, such occasions are very common: Multitasking and task-switching, dialogues and pop-ups, screen savers, advertisements and banners, notifications and reminders, radios and TVs, mobile and ordinary phones, and colleagues often divert attention to a message that is irrelevant from the point of view of the main task (McFarlane and Latorella 2002). Moreover, interruptions have an important role in work accidents and faulty thinking in work life (Saariluoma 2002). Consequently, interruptions provide an ecological situation to investigate the cooperation between the working memory systems in HCI-contexts.

By interruption we mean a discrete event during which attention is abruptly redirected to process information that is irrelevant to the ongoing main task. An interruption breaks the 'continuity' of processing of the main task and forces people to resume it later. Shifting attention to another task requires activation of task-related modules, inhibition of irrelevant modules and construction and maintenance of a new attentional set controlling the task (Allport *et al*. 1994, Rogers and Monsell 1995, Wylie and Allport 2000). From the LTWM point of view, one of the relevant differences between STWM and LTWM is the property of the latter that it cannot be easily interfered with secondary tasks. Therefore, if the person is to resume an interrupted task, task information must be saved to LTWM before the task-switch. How an interruption can disrupt this process, and in what reconditions disruption occurs, is the question investigated in this paper.

A brief review of experiments studying the interruption resistance of STWM vs. LTWM gives us some pointers to answering our question. The classic research into short-term remembering showed that information in STWM is in a very instable mode. Practically, any demanding interrupting task wipes off contents of STWM, causing permanent forgetting (Brown 1958, Peterson and Peterson 1959, Murdock 1962). However, experiments in chess and other domains showed that in skilled tasks interrupting secondary tasks do not cause essential impairment (Charness 1976, Frey and Adesman 1976, Lane and Robertson 1979, Chase and Ericsson 1981, Saariluoma 1991). Therefore, one has to assume that, in skilled tasks, information is transferred from short- to long-term memory at some point when keeping up an ongoing task during an interruption. Naturally, the information controlling human performance during longer interaction episodes must be stored into a more permanent storage than STWM, because any interrupting task would wipe off the representations needed to maintain the continuity of the main task.

Once stored to long-term working memory, information is not very vulnerable to secondary tasks. Indeed, neither the length nor the amount of memory load seems to have any substantial effect of interruption costs (Charness 1976, Frey and Adesman 1976, Gillie and Broadbent 1989). Sometimes one can even find reverse effects in which interruptions actually improve performance (Boltz 1992, Speier *et al*. 1997, 1999, Fisher 1998). However, even in skilled tasks, interruptions may impair the level of performance. Professionals often have poor strategies in managing interruptions in their work leading to negative effects on performance (Cohen 1980, Kirmeyer 1988, Eyrolle and Cellier 2000). Similarity between the main and interruption task is another factor causing interference in skilled tasks (Czerwinski *et al*. 1991), but not in all cases (Charness 1976, Glanzer *et al*. 1984). Experts in mental abacus, for example, are highly vulnerable to retroactive interference caused by attending to irrelevant abacus configurations and blindfolded chess masters to concurrent secondary tasks (Hatta *et al*. 1989, Saariluoma 1991).

What are the differences between skilled activities resistant to interruptions and those not vulnerable to them? One central assumption of the theory of LTWM is that interruption can disrupt transfer from STWM to LTWM. It is known that, when the interrupting task occurs *simultaneously* with the main task and thus disrupts transfer from short-term into longterm working memory, interruptions impair memory (Saariluoma 1991). This explanation can naturally also be supported by the vast research on modular working memory, in which simultaneous tasks substantially impair encoding performance (e.g. Baddeley and Hitch 1974, Baddeley 1986, Pashler *et al*. 2001). More specifically, if the attained level of encoding speed is not rapid enough, an abruptly occurring interruption is predicted to cause forgetting and susceptibility to intrusions from prior knowledge by preventing proper safeguarding – encoding of retrieval cues to stable retrieval structures – to LTWM.

The reality of this mechanism is here investigated in three experiments. The main goal of the empirical work presented here is to investigate whether LTWM could be a theoretical construct representing a real explanatory value in discussing interruptions in HCI. If this were the case, it would also allow making psychologically motivated design rationales and hypotheses, and encourage efforts to continue investigating the role of this important theoretical construct in human-computer interaction.

2. Experiment 1

The problem explored in Experiment 1 is whether the effects of interruption are a consequence of poor encoding. If one does not accept the idea that the locus of interruption's effects on memory is on the encoding side, one may also think that the interruption can overload STWM. Specifically, it has been suggested that STWM holds pointers to representations residing in LTWM (Gobet and Simon 1996, Gobet 1998). These hypotheses make different predictions in memory performance. In the case of STWM overloading, interruption would only cause forgetting, but not so intrusions. Directing attention to an interrupting task would override pointers and hamper later retrieval of information because, although the information would be intact in LTWM, it could not be easily retrieved without the pointers. Therefore, to investigate whether interruption's main effect is STWM overloading or poor encoding, we devised an experimental paradigm where interruption-caused intrusions and omissions (as indicators of 'pure' forgetting) were measured.

The experimental task in Experiment 1 is a laboratory equivalent of a situation where the processing of a message is interrupted by diverting to a competing message in the interface that must also be processed. A form of verbal representation was selected because of its central role in information technology. Participants shadowed (i.e. repeated aloud) a text spoken by a person on a video. The reader was shown on the screen and the speech presented through headphones. In this form of bimodality the visual channel (i.e. movement of lips) is by no means redundant but enhances comprehension of speech (McGurk and MacDonald 1976). Video was selected because it is also a relevant presentation medium for modern and future user interfaces (e.g. multimedia and multimodal interfaces). After one minute of shadowing, the presentation of the main text stopped. In the experimental condition (interruption), another text was shadowed for 30 s. In the control condition, a black screen appeared for 30 s. Immediately afterwards, the main (i.e. the first) text was recalled. Omissions and intrusions were analysed from verbal protocols.

2.1. *Method*

2.1.1. *Participants*: Twelve undergraduate students from an introductory course on cognitive science at the Open University of Helsinki were recruited as participants. The mean age of the participants was 30.5 years, ranging from 20 to 51 years (SD = 10.7). Nine of the participants were male, three of them female.

2.1.2. *Apparatus and materials*: Six main texts were prepared for the present experiment. The lengths of the main texts ranged from 95 to 105 words. Special care was taken to minimize the number of difficult words and complex sentence structures by asking a colleague to read and evaluate the texts. To reduce the build-up of proactive interference (and carry-over effects), topics for the texts were selected from different domains (Wickens *et al*. 1963). The contents and style of the texts resembled short news stories.

The topics of the main texts were: 'History of the Alcoholics Anonymous movement,' 'Effects of heat on outdoor work,' 'Information technology education in comprehensive schools in Finland,' 'Paintball,' 'Food-related allergies,' and 'The Saimaa Canal.' Six other texts were prepared as interrupting texts using the same procedure as with the main texts. Interrupting texts were approximately half as long (in words) as the main texts. Their topics were also selected from different domains. The topics of the interrupting texts were: 'Down Syndrome in mice,' 'Finnish mites,' 'Rose Champaign,' 'Krav Maga,' 'Students coming from neighbouring municipalities to senior high schools in Helsinki,' and 'Selecting an ergonomic desktop for computer work.' For an example of a main text and an interrupting text, consult Appendix A. All materials were in Finnish.

One person read out loud the main texts for the video while another person spoke the interrupting texts. The video was digitized and edited on a computer to form the stimulus material for the experiment. For a screen-shot, see figure 1.

All experiments were administered on an IBM-PC compatible computer with a 17- inch monitor. Videos were presented using *Windows Media Player* with an on-screen resolution of 300×400 pixels, spanning a 10×10 cm area on the display. Speech was presented through headphones. The sampling frequency of the audio speech was 32 KHz. Audio manipulation software (*CoolEdit*) was used to reduce background noise in the speech track.

2.1.3. *Procedure and design*: Participants were greeted and told that they would be participating in an experiment studying recall of computer-presented speech. They were then told that their task was to shadow everything they were going to hear and that they had to memorize and recall only the first text. It was then explained that a white cross on the display means that they will have to tell everything they remember about the first text, and that their recall will be tape-recorded. It was stressed that word-for-word recall was not necessary. No specific instructions were given concerning what to do during a pause. The whole procedure was practiced once before the experimental trials began.

A

B

Figure 1. Screenshot from Experiments 1, 2, and 3. Person A spoke the main texts and person B the interrupting texts.

In the experiment, participants shadowed the video for one minute, after which there was either a pause or an interruption for another 30 s. In the interruption condition, an interrupting text was presented, whereas in the control condition, a silent break occurred. Every participant underwent three interruption trials and three control trials. Interrupting text was always shadowed to ensure that participants attended to the stimuli. A white cross appeared on the screen to indicate the beginning of free recall. Free recall terminated when the participant stated that nothing more could be recalled (i.e. spontaneous ending of free recall; van Bergen 1968). The possibility to rest between the trials was provided.

The reader should notice that the experimental task used in this and subsequent studies contain a test of memory immediately after the interruption or pause. Also, it does not involve interactions other than shadowing, which is an easy, although attention-demanding, task. These properties of the experimental setup allow for a more direct indicator of the interruption's affects specifically on memory than does observing interaction after task resumption.

A within-subject design was used. The presentation order of the main texts (6) and the order of interruptions and pauses (2) were counterbalanced across participants by rotation, yielding a total of 12 combinations.

2.1.4. *Analysis and results*: Free recall protocols were analysed in three stages. First, following Kintsch and van Dijk (1978), texts were analysed to twelve constituent abstract (gist) propositions worth one point each. Some propositions consisted of two smaller micropropositions worth half a point. The maximum score (perfect recall) is 12 points. In the second stage, two outsiders independent of each other scored protocols according to the presence/absence of the propositions. Third, erroneous claims were classified as prior knowledge, interrupting text, or within-text intrusions. *Within-text intrusion* is defined as having its source in the text; for example, two dates mentioned in one of the texts were often mixed. *Interrupting text intrusion* is defined as having its source in the interrupting text and *prior knowledge intrusion* in none of the texts presented in the experiment. (Examples of intrusions in Experiment 3 are given in Appendix B.) The overall consensus between the two classifiers was 93%. Conflicting classifications were discussed and a common ground was established on the basis of the discussion.

Scores are presented in table 1. Mean recall accuracy in the control condition was 4.74 (SD = 1.73) and 4.00 (SD = 1.97) in the interrupted reading condition. A one-tailed paired samples t-test conducted on logarithmically transformed accuracy scores revealed a significant effect of interruption on recall accuracy, $t(11) = 3.049$, $p < .01$.[1] There were 0.67 (SD = 0.83) prior knowledge intrusions in the interrupted, vs. only 0.28 (SD = 0.51) in the control condition. This difference was significant: $t(11) = 3.084$, $p < 0.01$. However, there was no significant difference in the number of within-text intrusions (M = 0.56, SD = 0.65; M = 0.39, SD = 0.55, respectively, for interruption and control condition), $t(11) = 1.254$, *n.s.* No intrusions from the interrupting text were found from the protocols.

The reader should note that a normal free recall protocol has a number of (prior knowledge) intrusions arising from prior-knowledge-driven interpretative reconstructive processes carried out during both encoding and retrieval (Bartlett 1932, Alba and Hasher 1983). Hence, it is understandable that the absolute number of intrusions was relatively high.

2.2. *Discussion*

Two findings were made. Firstly, the interrupting message had a clear negative effect on memory accuracy in comparison to the silent pause. Recall accuracy was 16% worse after the interruption. No intrusions were found from the interrupting text, most likely because the interrupting text was semantically too different to be confused during retrieval. This is in agreement with our overall expectation that tasks effecting encoding may impair recall accuracy.

Secondly, there were significantly more intrusion errors in the interruption condition than in the control condition. The fact that intrusions were from prior-knowledge, not from the interrupting text, implies that the disruption of interruption takes place during the semantic elaboration of the main text. *Semantic elaboration* involves building meaningful associations, on the one hand, within chunks residing in working memory, and on the other hand, between these chunks and prior knowledge in long-term memory. When this process is disrupted by a requirement to abruptly direct attention to process irrelevant material, it seems that the to-be-remembered material cannot be properly safeguarded from proactive interference caused by prior knowledge. Organization of retrieval structures and quality of encoded retrieval cues is poorer when elaboration is disrupted. This kind of disruption causes an erroneous representation. These effects suggest that semantic elaboration is partially prevented by the interrupting text.

3. Experiment 2

The first goal of the second experiment is to replicate the results of the first experiment in a situation where four interrupting texts occurred instead of just one. The second goal is to provide a further test for the claim that the texts (or pointers to them) are stored in short-term working memory, not long-term working memory. One could assume that information transformation from short-term to long term working memory is not decisive but interruptions impair memory rather because the main task is stored in short-term working memory and the interrupting task simply causes short-term memory overflow. If STWM is the *only* storage used, four texts would clearly exceed its severely limited capacity and hamper comprehension of the texts. In contrast, long-term working memory is relatively robust to such an increase in memory load (Frey and Adesman 1976, Lane and Robertson 1979, Saariluoma 1991, Ericsson and Kintsch 1995). Therefore, if the STWM storage hypothesis is correct, we would expect to see a substantially lower level of performance than in Experiment 1. In contrast, if the LTWM storage hypothesis is correct, interruptions should only cause a slight impairment very close to the level seen in Experiment 1.

3.1. *Method*

3.1.1. *Participants*: Twelve participants were recruited from the same pool as in Experiment 1. The mean age of participants was 27.1 years (SD = 8.8), ranging from 21 to 48. Six of the participants were male, six female.

3.1.2. *Apparatus and materials*: Six main texts were adopted from Experiment 1 and two new main texts were prepared using the same procedure as in Experiment 1. The topics of the main texts were: 'History of the Alcoholics Anonymous movement,' 'Effects of heat on outdoor work,' 'Information technology education in comprehensive schools in Finland,' 'Paintball,' 'Food-related allergies,' 'The Saimaa Canal,' 'Purpose and tasks of the EU Commission,' and 'Anni Swan, a Finnish novelist.' Eight other texts were prepared for the interruption condition, six of which were already used in Experiment 1. The topics were: 'Down Syndrome in mice,' 'Finnish mites,' 'Rose Champaign,' 'Krav Maga,' 'Students coming from neighbouring municipalities to senior high schools in Helsinki,' 'Practicing self-defence

Table 1. Recall accuracy and intrusions in Experiment 1.

		Intrusion source	
Condition	Accuracy	Within-text	Prior knowledge
Interruption	4.00 (1.97)	0.56 (0.65)	0.67 (0.83)
Pause	4.74 (1.73)	0.39 (0.55)	0.28 (0.51)

Note: Standard deviations are in parentheses.

situations,' 'Conflicts between municipalities in the Helsinki region,' and 'Selecting an ergonomic desktop for computer work.'

Two sets of videos were prepared for the experiment. Both sets had four main texts and four pauses and four interrupting texts. Videos were prepared as in Experiment 1.

3.1.3. *Procedure and design*: Participants were told that they would be participating in an experiment studying the recall of computer-presented speech, and that their task was to shadow everything they heard but to memorize only the texts spoken by the person in a striped shirt (see figure 1A), and that 30 s after they had heard altogether four texts spoken by the striped-shirt person, a white cross would show on the display and the experimenter would write on a piece of paper the heading of the main text which was to be recalled. It was stressed that even though only the text spoken by the striped-shirt person should be remembered, all speech should be shadowed. The whole procedure was then practised once before the experiment. Each participant underwent two trials and had an opportunity to rest between them.

A within-subject design was used. Counterbalancing was conducted by rotating the order of experimental conditions (two) and videos (two) and the to-be-recalled main texts (four) across participants.

3.1.4. *Analysis and results*: Recall protocols were analysed using the same procedure as in Experiment 1. Consensus between the two classifiers was 93%.

Consult table 2 for scores. Mean recall accuracy for non-interrupted texts was 4.34 (SD = 2.05), and for interrupted 3.97 (SD = 2.79). A paired-samples *t*-test (one-tailed) for reciprocally transformed accuracy scores yielded a significant main effect of reading condition for recall accuracy, $t(15) = 2.129$, $p < 0.05$.[2] Since no interrupting text intrusions were found, *t*-tests were conducted only for prior knowledge and within-text intrusions. No significant differences between the conditions were found for either of these two dependent variables, both ts < 0.824, *n.s.*

3.2. *Discussion*

Interruption again had a negative effect on recall accuracy. Recall accuracy was 8.5% lower after an interrupting text than after a pause. Although not significant, the pattern in intrusions also followed the same trend: there were slightly more intrusions in the interruption condition. The results of the first experiment were thus partly replicated in this experiment, which suggests that the main locus of memory disruption caused by the interruption is more in the semantic elaboration of memory trace. These results thus support the long-term working memory storage hypothesis.

In contrast, the findings do not support the short-term working memory hypothesis, according to which STWM holds pointers referring to representations in LTWM over interrupting tasks. The requirement for keeping pointers to four texts (each involving 12 gist propositions) over four interruption periods of 30 s clearly exceeds the supposed four chunk capacity of STWM (Baddeley 1986, Cowan 2001).

4. Experiment 3

The two previous experiments suggest that the negative effect of an interrupting message on semantic elaboration may partly explain the impaired performance. The purpose of the final experiment was to investigate a situation where the interrupting message is semantically related to the main message. Supposing that interruptions impaired *exclusively* semantic elaboration, semantic-relatedness should have no effect on performance. However, another mechanism is supposed by the LTWM theory: *retroactive interference*. Retroactive interference refers to response competition *during retrieval*. For adept readers, this would occur only when retrieval structures in LTWM are not properly safeguarded. For adept readers, this would only happen when the texts are semantically close enough. In order to safeguard representations from semantic interference, effort is needed in strengthening and building associations within a representation, strengthening distinctive features of each representation, and separating features that are similar between representations. If these operations cannot occur, the LTWM theory would predict interference upon retrieval.

As in Experiment 1, a text was shadowed. After one minute, shadowing was interrupted by another text for 30 s. In the control condition, the interrupting text was from another domain. In the first experimental condition, it was a continuation of the main text. In the second experimental condition, it was partially overlapping with the main text.

4.1. *Method*

4.1.1. *Participants*: Eighteen participants were recruited for Experiment 3. Ten of them were students from an introductory course in cognitive science arranged by the University of Helsinki. Eight participants were students of computer science who volunteered for

Table 2. Recall accuracy and intrusions in Experiment 2.

		Intrusion source	
Condition	Accuracy	Within-text	Prior knowledge
Interruption	3.97 (2.79)	0.44 (.73)	0.44 (0.63)
Pause	4.34 (2.05)	0.25 (.58)	0.50 (0.63)

Note: Standard deviations are in parentheses.

the experiment. The mean age of participants was 24.5 years (SD = 3.87), ranging from 19 to 28. Ten of the participants were male, eight of them female.

4.1.2. *Apparatus and materials*: The main texts were adopted from Experiment 1. The topics were: 'Anni Swan, a Finnish novelist,' 'History of the Alcoholics Anonymous movement,' 'Effects of heat on outdoor work,' and 'The Saimaa Canal.' For the control condition, *interrupting texts from another domain* (than the main text) were used. These texts were taken from Experiment 1. New interrupting texts were prepared for the two experimental conditions. *Continuation of a main text* was simply defined as continuing the main text from where it ended. *Partially overlapping text* was defined as having propositions concerning the same issues as in the related main text. Two outsiders were asked to classify the constructed interrupting texts using these definitions. Consensus between the two was 100%. An example of a main text and interrupting texts is given in Appendix A. Main and interrupting texts in the second and third condition were spoken on the video by the same person (in a striped shirt, figure 1A). In contrast, another person spoke the interrupting text from another domain (figure 1B). We added a visual sign (white cross) between the main and the interrupting text to indicate task change.

4.1.3. *Procedure and design*: Apart from a few changes, the instructions, practice, and procedure were the same as in Experiment 1. First, participants were told that a visual sign (a white cross) indicated a switch from the main (the to-be-remembered text) to another text that should also be shadowed. Second, they were instructed that the end of the video (black screen) would mark the beginning of the free recall.

A within-subject design was used. As in previous experiments, counterbalancing was carried out by rotation.

4.1.4. *Analysis and results*: Recall protocols were analysed using the same procedure as in the previous experiments. Consensus between the two classifiers was 96%.

Mean accuracy for recall after an interrupting text from another domain was 3.74 (SD = 1.92), 3.04 (SD = 2.14) after interrupting text continuing the main text, and 3.25 (SD = 1.48) after partially overlapping interrupting text. A paired-samples t-test for logarithmically transformed scores yielded a significant difference between interrupting text from another domain and interrupting text that continued the main text, $t(17) = 2.141$, $p < 0.05$, as expected. However, the difference between control condition and overlapping condition was not significant, $t < 0.908$. Consult table 3 for scores.

No significant differences were found in the number of within-text or prior knowledge intrusions, both ts < 1.342. There were, on average, 0.25 and 0.11 intrusions from the interrupting texts in the continuation of the main text and the semantically overlapping condition, respectively. There were no intrusions from the interrupting text in the control condition where the interrupting text was from another domain. The difference between two conditions, the control and the continuation conditions, was significant, $t(17) = 3.000$, $p < 0.01$, while between the control and the overlapping conditions, and between the continuation and the overlapping conditions, it was not, both ts < 1.158. Two examples of recall protocols with intrusions from interrupting texts are given in Appendix B.

4.2. *Discussion*

Experiment 3 shows that a semantically similar interrupting message can disrupt remembering more than a message from another domain. There were significantly more omissions when the interrupting text was a continuation of the main text in comparison to the situation where it was from another semantic domain (control). However, there were no differences in the number of prior knowledge intrusions or within-text intrusions. Instead, intrusions from the interrupting text were found in the two conditions where interrupting texts were semantically close to main texts. This finding is consistent with the hypothesis that semantic similarity of the interrupting message causes interference that disrupts retrieval of information from LTWM. The locus of this interference is presumably at retrieval where several similar strong-but-wrong responses match the retrieval criteria and it is not easy to distinguish between them (Ericsson and Kintsch 1995). When the interrupting message appears closely after the main message, the representations are not easily distinguished using temporal cues. Moreover, the modality and presentation format of the interrupting message was identical to the main message, eliminating the possibility of using perceptual cues or source memory in retrieval. In this case, the semantic relatedness of the interrupting message easily causes intrusions and confusions upon retrieval, a phenomenon we call here semantic interference.

Moreover, it is likely that the two interruption-induced mechanisms proposed here interact. When semantic elaboration is interrupted, as we showed in Experiments 1 and 2, the representation is poorly integrated in LTWM and thus more susceptible to all forms of interference, including that caused by the interrupting message. Thus, abrupt interruptions disrupting elaboration would cause increased vulnerability to semantic interference as well. This prediction is examined in further studies by the authors.

5. General discussion

Long-term working memory constructs have received relatively little attention, so far, in human-computer interaction research. A possible reason is that much of recent HCI research has concentrated on usability rather than on psychological constructs. Our experiments were designed to investigate the possible role of long-term working memory in remembering important task information over interruptions in human-computer interaction.

In this paper we focused on interruptions because they are evidently relevant in investigating the functions of the two working memory systems in human computer interaction. The reason for this is the vulnerability of information in short-term working memory and the relative invulnerability of information in long-term working memory for concurrent and subsequent secondary tasks. This means that interruptions should affect the former type of information but not to the latter. Therefore, we decided to examine the locus of impairment from encoding rather than from storage.

Our experiments illustrated a vulnerability to interruptions; in all three experiments a negative effect of interruption was shown. In the two first experiments, evidence indicated that one locus of this effect is on the encoding of information. Intrusions were found from related prior knowledge in protocols of interrupted trials, indicating poor safeguarding during encoding. We concluded that the main effect of interruption takes place when information is transferred by semantic elaboration from short-term to long-term working memory. The disturbed semantic elaboration involves building meaningful within-chunk associations, and associations between chunks and prior knowledge in long-term memory.

In Experiment 3, we found systematic intrusions from semantically similar interrupting messages. This suggests that the interrupting task, if semantically close enough, can compete with the main task upon retrieval, providing evidence for another mechanism postulated by the LTWM theory, semantic (retroactive) interference. The focus of this mechanism is in the retrieval stage. Increasing semantic distance between the main task and interrupting task prevented confusions and improved recall.

6. Implications for human – computer interaction

The extent to which memory for information content is reliable, trustworthy, and accurate is crucial in the information age. The question of how interaction with information contents, not just with the user interface they are wrapped in, should be designed in order to

Table 3. Recall accuracy and intrusions in Experiment 3.

		Intrusion source		
Condition	Accuracy	Within-text	Prior knowledge	Interrupting text
Interrupting text from another domain	3.74 (1.92)	0.17 (0.34)	0.67 (0.66)	–
Interrupting text as a continuation of main text	3.04 (2.14)	0.15 (0.30)	0.42 (0.43)	0.25 (0.35)
Interrupting text semantically overlapping with main text	3.25 (1.48)	0.14 (0.28)	0.61 (0.61)	0.11 (0.27)

Note. Standard deviations are in parentheses.

enhance comprehension and remembering has not been investigated enough. As we have shown, the theory of LTWM provides a serious starting point for this kind of research.

We showed that interruptions cause not only forgetting (i.e. omissions) but also distortions (i.e. intrusions) of main task representations. Thus, interruptions not only force users to restudy the main task, resulting in frustration and loss of time, but several other negative consequences can occur due to erroneous representations: Interrupted users could search for information in the wrong place, at the wrong time, and with the wrong methods; moreover, interrupted users may end up with faulty, biased, or distorted conceptions of the material they are reading. Furthermore, once acquired, misconceptions may persist for a long time, since users are typically too reluctant or lazy to update representations even though the necessary information is readily available in the user interface (Gray and Fu 2001). Carelessly designed interrupting messages are thus likely to negatively affect all components of usability: efficiency, effectiveness, and satisfaction (ISO-9241-11 definition of usability).

However, apart from usability and human – computer interaction, task interruptions are important also in the more general context of human – technology interaction. Indeed, interrupting messages play a special role in real time systems in office, industrial, traffic, and aviation settings. For example, in a fatal aviation accident (NTSB 1988), an interruption of a pilot during a checklist led to a critical omission of an item in the list (deployment of flaps/slats to retract position during takeoff). In power plant incidents, one study shows that 25% of shut-down incidents are partly accounted for by interruptions of primary tasks (Griffon-Fouco and Ghertman 1984). Similarly, in the work of mobile office workers, 41% of interruptions, which occurred on average four times in an hour, resulted in discontinuing the interrupted task (O'Conaill and Frohlich 1995).

Our most basic suggestion for practitioners is that they should minimize the number of uncontrolled interruptions during attention to information content (see also: Gillie and Broadbent 1989, Detweiler *et al.* 1994), at least when the main content is of some importance. Moreover, if possible, the time spent on processing competing material should be minimized because virtually all processing during an interruption may take part in building interference between the tasks (Nairne 2002). Our more specific suggestions stem from the nature of the two processes that interrupting messages are proposed to disrupt.

The idea that interrupting messages may disrupt semantic elaboration has interesting consequences to design rationale. It implies that main messages should be organized into small and coherent chunks (e.g. in the case of textual content) or episodes (e.g. in the case of narratives, videos, animation etc.). Making information chunks and episodes small and coherent should minimize the possibility of an interruption occurring during elaboration. For example, if WWW material is carelessly divided to span several separate pages, diversions of attention to competing messages can have deleterious effects. In such situations, short summaries or restatement sentences in the beginning of a new page or chapter may serve to help reorientation.

Moreover, users should be given some time for making the task-switch to the interrupting message. This would give time to complete the elaboration. Disruption occurring in the midst of elaborating, as we have shown, affects memory accuracy negatively especially in such tasks where the user does not possess memory skills that ensure rapid encoding to persistent and robust retrieval structures in LTWM. A pause, by contrast, occurring between coherent chunks or episodes can even enhance learning and remembering (Boltz 1992), most likely because it increases the temporal distinctiveness of the memory traces (see Baddeley 1986, for a review of the passive recency effect), while causing only minimal disturbance to semantic elaboration and only little semantic interference, unlike interrupting messages that require semantic orientation. Similarly, Czerwinski *et al.* (2000) have suggested that interruptions should be timed in between tasks, not in the middle of them. Dividing material into small entities gives more freedom for the user to voluntarily initiate a break or a pause in between episodes or chunks.

In addition, the disturbance of semantic elaboration is also a natural explanation for why interruption during high memory load has often been demonstrated to be more disruptive than during low memory load (Gillie and Broadbent 1989, Speier *et al.* 1999; see also Detweiler *et al.* 1994, Latorella, 1999, McFarlane 1999).

Deriving from the semantic interference hypothesis, we recommend that the interrupting message should be semantically distinctive (see also Czerwinski *et al.* 1991). It must be noted that according to LTWM, mere *perceptual* dissimilarity between the messages may not suffice to eliminate interference, since texts are encoded primarily based on the semantic gist, whereas perceptual detail can be quickly forgotten. Moreover, the way we retrieve texts is semantic in nature, and does not necessarily tap perceptual components of the trace to the same extent. In tasks other than reading where representations are not primarily semantic in nature, or where experienced users have learned to associate the perceptual cues provided by the UI with the semantics in the retrieval structure, perceptual distinctiveness may suffice (see Rhodes *et al.* 2000).

Finally, we want to note that interruptions are a topic worth investigating. It has practical implications and it should be taken into account in interaction design. Interruption design is a part of designing effective interfaces. We should not underestimate the importance of interruptions in explaining faulty thinking, lapses of memory, and attentional biases in work life. Indeed, interruptions are a problem with interesting practical perspectives.

Notes

1. Logarithmic transformation is a preprocedure for normalizing positively skewed response data. Response data in experiments with natural stimuli (such as expository texts used here) are often skewed because of large differences between participants in the amount and quality of prior knowledge.

2. Reciprocal transformation is used to stabilize a notable increase in variance above a certain threshold. Again, the need for transformation was caused by large individual differences in material-related expertise.

Acknowledgements

We thank Pauli Salo and Jukka Purma for their co-operation in recruiting participants, Jyrki Oraskari and Tero Haahtela for helping with materials, and B. Bailey, D. McFarlane, and an anonymous referee for their valuable comments on the first version of the manuscript. Parts of this work were done for the first author's Master's Thesis in cognitive science at the University of Helsinki.

References

Alba, J. W., and Hasher, L. 1983, Is memory schematic? *Psychological Bulletin*, **93,** 203 – 231.

Allport, A., Styles, E. A., and Hsie S. 1994, Shifting attentional set: Exploring the dynamic control of tasks. In C. Umiltà and M. Moskovitch (eds) *Attention and Performance XIV: Conscious and nonconscious processes*, (Cambridge, MA: MIT Press), pp. 421 – 452.

Baddeley, A. 1986, *Working Memory* (Cambridge, UK: Cambridge University Press).

Baddeley, A. 1996, Exploring the central executive. *The Quarterly Journal of Experimental Psychology*, **49A**(1), 5 – 28.

Baddeley, A. 2000, The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences*, **4,11,** 417 – 423.

Baddeley, A., and Hitch, G. 1974, Working memory. In G. H. Bower (ed) *The psychology of learning and motivation: Advances in research and theory*, Vol. 8, (New York: Academic Press), pp. 47 – 89.

Bartlett, F. C. 1932, *Remembering: A Study in Experimental and Social Psychology* (New York: Cambridge University Press).

Boltz, M. 1992, The incidental learning and remembering of event durations. In F. Macar, V., and W.J. Friedman (eds) *Time, Action and Cognition: Towards Bridging the Gap*, (Kluwer Academic Publishers, NATO ASI Series), pp. 153 – 163.

Brown, J. 1958, Some tests of the decay theory of immediate memory. *Quarterly Journal of Experimental Psychology*, **10,** 12 – 21.

Charness, N. 1976, Memory for chess positions: resistance to interference. *Journal of Experimental Psychology: Human Learning and Memory*, **2,** 641 – 653.

Chase, W. G., and Ericsson, K. A. 1981, Skilled memory. In J.R. Anderson (ed) *Cognitive skills and their acquisition*, (Erlbaum).

Cohen, S. 1980, Aftereffects of stress on human performance and social behavior: A review of research and theory. *Psychological Bulletin*, **88**(1), 82 – 108.

Cowan, N. 2001, The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral & Brain Sciences. Special Issue*, **24**(1), 87 – 185.

Czerwinski, M. P., Chrisman, S. E., and Rudisill, M. 1991, Interruptions in multitasking situations: The effects of similarity and warning (NASA Technical Report #JSC-24757).

Czerwinski, M. P., Cutrell, E., and Horvitz, E. 2000, Instant messaging and interruption: Influence of task type on performance. Proceedings of OZCHI 2000, December 2000.

Detweiler, M. C., Hess, S. M., and Phelps, M. P. 1994, Interruptions and workng memory. Technical Report, University Park, CA: The Pennsylvania State University.

Ericsson, K. A., and Delaney, P. F. 1999, Long-term working memory as an alternative to capacity models of working memory in everyday skilled performance. In A. Miayke, and P. Shah (eds) *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control*, (Cambridge, UK: Cambridge University Press), pp.257 – 297.

Ericsson, K. A., and Kintsch, W. 1995, Long-term working memory. *Psychological Review*, **102,** 211 – 245.

Ericsson, K. A., and Lehmann, A. C. 1996, Expert and exceptional performance: Evidence for maximal adaptation to task constraints. *Annual Review of Psychology*, **47,** 273 – 305.

Eyrolle, H., and Cellier, C. 2000, The effects of interruptions in work activity: field and laboratory results. *Applied Ergonomics*, **31,** 537 – 543.

Fisher, C. D. 1998, Effects of external and internal interruptions on boredom at work: two studies. *Journal of Organizational Behavior*, **19,** 503 – 522.

Frey, P. W., and Adesman, P. 1976, Recall memory for visually presented chess positions. *Memory and Cognition*, **4,** 541 – 547.

Gillie, T., and Broadbent, D. 1989, What makes interruptions disruptive? A study of length, similarity, and complexity. *Psychological Research*, **50,** 243 – 250.

Glanzer, M., Fischer, B., and Dorfman, D. 1984, Short-term storage in reading. *Journal of Verbal Learning and Verbal Behavior*, **23,** 467 – 486.

GOBET, F. 1998, Expert memory: A comparison of four theories. *Cognition*, **66**(2), 115–152.
GOBET, F., and SIMON, H. A. 1996, Templates in chess memory: A mechanism for recalling several boards. *Cognitive Psychology*, **31**(1), 1–40.
GRAY, W. D., and FU, W.-t. 2001, Ignoring perfect knowledge in-the-world for imperfect knowledge in-the-head: Implications of rational analysis for interface design. *CHI Letters*, **3**(1), 112–119.
GRIFFON-FOUCO, M. and GHERTMAN, F. 1984, Receueil de donnees sur les facteurs humains a Electricite de France. In *Operational Safety of Nuclear Power Plants*, (Vienna: International Atomic Energy Agency), pp. 157–172.
HATTA, T., HIROSE, T., IKEDA, K., and FUKUHARA, H. 1989, Digit memory of Soroban experts: Evidence of utilization of mental imagery. *Applied Cognitive Psychology*, **3,** 23–33.
KINTSCH, W., and VAN DIJK, T. A. 1978, Towards a model of text comprehension and production. *Psychological Review*, **85,** 363–394.
KIRMEYER, S. L. 1988, Coping with competing demands: Interruptions and the type A pattern. *Journal of Applied Psychology*, **73**(4), 621–629.
LANE, D. M., and ROBERTSON, L. 1979, The generality of levels of processing hypothesis: An application to memory for chess positions. *Memory and Cognition*, **7,** 253–56.
LATORELLA, K. A. 1999, Investigating interruptions: Implications for flightdeck performance. NASA/TM-1999-209707, National Aeronautics and Space Administration, Hampton.
MCFARLANE, D. C. 1999, Coordinating the interruption of people in humancomputer interaction, In M.A. Sasseand and C. Johnson (eds) *Human-Computer Interaction – INTERACT'99*, (The Netherlands, IOS Press, Inc), pp. 295–303.
MCFARLANE D. C., and LATORELLA K. A. 2002, The scope and importance of human interruption in human-computer interaction design. *Human-Computer Interaction*, **17**(1), 1–61.
MCGURK, H., and MACDONALD, J. W. 1976, Hearing lips and seeing voices. *Nature*, **264,** 746–748.
MURDOCK, B. B. Jr., 1962, The serial position effect in free recall. *Journal of Experimental Psychology*, **64,** 482–488.
NAIRNE, J. S. 2002, Remembering over the short-term: The case against the standard model. *Annual Review of Psychology*, **53,** 53–81.
NTSB, 1988, *Aircraft accident report* N312RC (Washington DC: National Traffic Safety Board).
O'CONAILL, B., and FROHLICH D. 1995, Timespace in the workplace: Dealing with interruptions. Proceedings of CHI' 95 (New York, NY: ACM Press), pp. 262–263.
PASHLER, H., JOHNSTON, J., and RUTHRUFF, E. 2001, Attention and performance. *Annual Review of Psychology*, **52,** 629–651.
PETERSON, L. R., and PETERSON, M. J. 1959, Short term retention of individual verbal items. *Journal of Experimental Psychology*, **58,** 193–198.
RHODES, J. S., BENOIT, G. E., and PAYNE, D. G. 2000, Factors affecting memory for dynamically changing system parameters: Implications for interface design. Proceedings of the IEA 2000/HFES 2000 Congress (San Diego, CA: Human Factors and Ergonomics Society), pp. 284–285.
RICHMAN, H. B., STASZEWSKI, J. J., and SIMON, H. A. 1995, Simulation of expert memory using EPAM IV. *Psychological Review*, **102**(2), 305–330.
ROGERS, R. D.and MONSELL, S. 1995, Costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General*, **124,** 207–231.
SAARILUOMA, P. 1991, Aspects of skilled imagery in blindfold chess. *Acta Psychologica*, **77**(1), 65–89.
SAARILUOMA, P. 1995, *Chess players' thinking: a cognitive psychological approach* (London: Routledge).
SAARILUOMA, P., and KALAKOSKI, V. 1997, Skilled imagery and long-term working memory. *American Journal of Psychology*, **110**(2), 177–201.
SAARILUOMA, P., and KALAKOSKI, V. 1998, Apperception and imagery in blindfold chess. *Memory*, **6**(1), 67–90.
SAARILUOMA, P. 2002, *Ajattelu työelämässä* (Thinking in work-life). (Porvoo, Finland: WSOY).
SPEIER, C., VALACICH, J. S., and VESSEY, I. 1997, The effects of task interruption and information presentation on individual decision making. Proceedings of the International Conference on Information Systems, pp. 21–35.
SPEIER, C., VALACICH, J. S., and VESSEY, I. 1999, The influence of task interruption on individual decision making: An information overload perspective. *Decision Sciences*, **30**(2), 337–360.
VAN BERGEN, A. 1968, *Task Interruption* (Amsterdam, Netherlands: North-Holland).
WICKENS, C. D., BORN, D. G., and ALLEN, C. K. 1963, Proactive inhibition and item similarity in short-term memory. *Journal of Verbal Learning and Verbal Behavior*, **2,** 440–445.
WYLIE, G., and ALLPORT, A. 2000, Task switching and the measurement of 'switch costs'. *Psychological Research*, **63,** 212–233.

Appendix A

Example of main and interrupting texts. (A) is a main text used in all experiments and (B) the related interrupting text from another domain (Exps 1 and 2, and condition 1 in Exp 3). (C) and (D) are semantically similar interrupting texts from Experiment 3. Texts are translated from Finnish.

(A) Main Text

Effects of heat on outdoor work. Outdoor work during heat is particularly demanding. Consequences may be serious if motivation, alertness, or work capability decrease in a task that requires accuracy. According to the work safety regulations, even light work should have several pauses when the temperature rises over 28°C. Cases of heat-related sickness and illness are quite rare in Finland. However, sunstroke, fainting, exhaustion, or apoplexy may come as a surprise for any worker regardless of his physical condition. First symptoms include, e.g. decrease in work performance, feverish state, dizziness, and nausea. The most troublesome work places have both high temperature and high humidity. Air movement and warmth radiation also affect workers. Firemen, construction workers, asphalt workers, and machine operators are particularly at risk.

(continues...)

(B) Interrupting text from another domain

Finnish mites. A mite is round and flat in shape. The species belongs to a general group of spider animals. Finnish mites usually live as long as 2 to 4 years. Every stage of development in the mite's life requires a blood meal. Also, female mites need a blood meal in order to become fertile and to lay eggs. In the following, we discuss the characteristics of a young mite, larva.

(C) Interrupting text continuing the main text

Researchers at the Department of Occupational Health have experimentally investigated the effects of warmth on performance. Some of the participants have reported being just fine, but have fainted almost immediately afterwards. The number of work accidents is highest during June and July, which may be explained by the high number of incompetent summer employees. However, experimental studies suggest that heat may also play a role in the figure.

(D) Semantically overlapping interrupting text

Monitoring temperature is important for workers with cardiovascular and respiratory disorders. Workers in poor physical condition and high-aged workers are easily exhausted. Lack of sleep and overuse of alcohol also have negative consequences on heat resistance. Despite that, the most vulnerable group is those over-motivated workers who constantly underestimate their need for rest.

Appendix B

Two recalls with intrusions from interrupting text in Experiment 3. The related main text and interrupting text are presented in Appendix A. Intrusions from interrupting text are in italics. Protocols are translated from Finnish.

'The text told about how heat affects work. Yes, there must be pauses in work when the temperature rises high enough. Even though the person might be in good condition, heat may result in negative consequences: fainting etcetera. *Then it was said that the symptoms might appear without the person being able to monitor them. He might say that he is ok, but in the next moment he would faint due to heat*. In addition there's ... Well, that's it.'

'Apoplexy caused by heat... Heat during summers may cause troubles like apoplexy. Symptoms also include nausea and fever. *In Finland, due to the cold climate, effects of incompetent summer employees are more important than those of heat*. Moreover, according to some law or something, even light work should have pauses when the temperature is around 28 degrees.'